UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Convolutional Neural Network for Accurate Crowd Counting and Density Estimation

by Saeed Amirgholipour Kasmani

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy
under the supervision of Professor Xiangjian (Sean) He

University of Technology Sydney
Faculty of Engineering and IT

February 2021

Certificate of Original Authorship Template

Graduate research students are required to make a declaration of original authorship when they submit the thesis for examination and in the final bound copies. Please note, the Research Training Program (RTP) statement is for all students. The Certificate of Original Authorship must be placed within the thesis, immediately after the thesis title page.

Required wording for the certificate of original authorship

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, ***Saeed Amirgholipour Kasmani*** declare that this thesis, is submitted in fulfilment of the requirements for the award of ***PhD***, in the ***SEDE/FEIT*** at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.
**If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).*

**If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).*

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 26/02/2021

Collaborative doctoral research degree statement

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with ***CSIRO\Data 61***.

Indigenous Cultural and Intellectual Property (ICIP) statement

This thesis includes Indigenous Cultural and Intellectual Property (ICIP) belonging to *[insert relevant language, tribal or nation group(s) or communities]*, custodians or traditional owners. Where I have used ICIP, I have followed the relevant protocols and consulted with appropriate Indigenous people/communities about its inclusion in my thesis. ICIP rights are Indigenous heritage and will always remain with these groups. To use, adapt or reference the ICIP contained in this work, you will need to consult with the relevant Indigenous groups and follow cultural protocols.

ABSTRACT

Convolutional Neural Network for Accurate Crowd Counting and Density Estimation

by

Saeed Amirgholipour Kasmani

Nowadays, crowd and object counting has become an important task for a variety of applications, such as traffic control, public safety, urban planning, and video surveillance. It has also become a crucial part of building a high-level monitoring system such as video surveillance and crowd analysis. In these cases, dynamic crowd monitoring and analysis is extremely important for control management and social safety.

Like the other computer vision issues, crowd counting and density estimation come with various kinds of challenges such as high clutters, occlusions, non-uniform distributions of objects or people, and intra-scene and inter-scene variations in appearance. Researchers and industrial partners have attempted to design and develop many sophisticated models to address various issues that exist in crowd counting. Especially in recent years, the number of researches in the crowd counting era became overwhelming with the domination of deep-learning and Convolution Neural Networks (CNNs) based models in various computer vision tasks. In this thesis, we revisit the crowd counting and propose various novel solutions to this problem.

At first, we propose an Adaptive Counting Convolutional Neural Network (A-CCNN) and consider the scale variation of objects in a frame adaptively to improve the accuracy of counting. Our method takes advantages of contextual information to provide more accurate and adaptive density maps and crowd counting in a scene. Then, we focus on CNN pruning to further enhance the crowd counting models for real-time application and increase the performance of CCNN model. Thus, a new

pruning strategy is proposed by considering the contributions of various filters to the final result. The filters in the original CCNN model are grouped into positive, negative, and irrelevant types. We prune the irrelevant filters, of which feature maps contain little information, and the negative filters determined by a mask learned from the training dataset. Our solution improves the results of the counting model without fine-tuning or retraining the pruned model. Finally, we propose a novel Pyramid Density-Aware Attention-based network, abbreviated as PDANet, which leverages the attention, pyramid scale feature and two branch decoder modules for density-aware crowd counting. The PDANet utilises these modules to extract different scale features, focus on the relevant information, and suppress the misleading ones. Extensive evaluations conducted on the challenging benchmark datasets well demonstrate the superior performance of the proposed models in terms of the accuracy of counting as well as generated density maps over the well-known state-of-the-art approaches.

Dissertation directed by Professor Xiangjian (Sean) He
School of Electrical and Data Engineering

Dedication

I owe thanks to a very special person, my wife, Mozhgan for her continued and unfailing love, support and understanding during my pursuit of Ph.D degree that made the completion of thesis possible. This thesis work is dedicated to you who were always around at times I thought that it was impossible to continue, and you helped me to keep things in perspective. I greatly value her contribution and deeply appreciate her belief in me.

This work is also dedicated to my mother, Soraya, who had always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

Acknowledgements

At this moment of accomplishment, I would like to express my deep appreciation to my wonderful supervisor, Prof Sean He, who has been guiding, supporting, and encouraging me throughout my PhD candidature journey. He gave me care and love like a father, which I can never forget. I am also greatly indebted to my co-supervisor, Dr Wenjing Jia, for her involvement and guidance throughout my PhD. She offered me her positive appreciations, constructive criticisms, and valuable advices during my PhD research which helped me tackle the challenges and complete the research work. I extend thanks to Prof Dadong Wang, my co-supervisor from CSIRO/DATA 61, for his time, encouragement, and expertise throughout this journey. I was able to accomplish this research due to the tremendous and unforgettable supports from my supervisory panel.

I need to express my deep gratitude to UTS and CSIRO/ Data61 to sponsor support and scholarships. I am thankful to the University of Technology Sydney, to give me permission to conduct PhD studies and provide me with state-of-the-art facilities to do my experiments. I am thankful to all staff at the School of Electrical and Data Engineering, FEIT, and GRS staffs of UTS for their kind assistance and support. It is my fortune to acknowledge my friends' support throughout the research tenure gratefully.

In the end, my wife, my family, and my parents, I sincerely thank you for believing me and showing faith in me. I would not have been able to focus on my research without the ongoing support and unconditional love of my wife.

Saeed Amirgholipour Kasmani

Sydney, Australia, 2021.

List of Publications

Journal Papers

J-1. **Amirgholipour**, S., Jia, W., Liu, L., Fan, X., Wang, D. and He, X., 2021. PDANet: Pyramid Density-aware Attention Net for Accurate Crowd Counting. Neurocomputing, 451, pp.215-230.

J-2. Liu, L., Jia, W., Jiang, J., **Amirgholipour**, S., Wang, Y., Zeibots, M. and He, X., 2020. Denet: A universal network for counting crowd with varying densities and scales. IEEE Transactions on Multimedia, 23, pp.1060-1068.

J-3. Liu, L., **Amirgholipour**, S., Jiang, J., Jia, W., Zeibots, M. and He, X., 2019. Performance-enhancing network pruning for crowd counting. Neurocomputing, 360, pp.246-253.

Conference Papers

C-1. **Amirgholipour**, S., He, X., Jia, W., Wang, D. and Zeibots, M., 2018, October. A-CCNN: Adaptive CCNN for density estimation and crowd counting. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 948-952). IEEE.

Contents

List of Figures

Abbreviation

CNN - Convolutional Neural Network

MAE - Mean Absolute Error

RMSE - Root Means Square error

Weak-Sup: Weak Supervised

Self-Sup: Self Supervised

Semi-Sup: Semi Supervised

Un-Sup: Un Supervised.

Fully-Sup: Fully Supervised

FIb: Full Image-Based

Pb: Patch-based

DOF: Degree Of Freedom

HPs: Hyper Parameters

Nomenclature and Notation

Capital letters denote matrices.

Lower-case alphabets denote column vectors.

$(.)^T$ denotes the transpose operation.

I_n is the identity matrix of dimension $n \times n$.

0_n is the zero matrix of dimension $n \times n$.

$\mathbb{R}$, $\mathbb{R}^+$ denote the field of real numbers, and the set of positive reals, respectively.